U.S.N.A --- Trident Scholar Project Report; no. 376 (2008)

EVOLUTION IN A TEST TUBE
EXPLORING THE STRUCTURE AND FUNCTION OF RNA PROBES

by

Midshipman 1/c Jeffrey E. Vandenengel
United States Naval Academy
Annapolis, Maryland

Certification of Adviser Approval

Assistant Professor Daniel P. Morse
Chemistry Department

Acceptance for the Trident Scholar Committee

Professor Joyce E. Shade
Deputy Director of Research and Scholarship

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATE COVERED
	2 May 2008	

4. TITLE AND SUBTITLE

Evolution in a Test Tube: Exploring the Structure and Function of RNA Probes

5. FUNDING NUMBERS

6. AUTHOR(S)

Vandenengel, Jeffrey E.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)		8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING/MONITORING AGENCY REPORT NUMBER
US Naval Academy Annapolis, MD 21402		**Trident Scholar project report no. 376 (2008)**

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
This document has been approved for public release; its distribution is UNLIMITED.	

13. ABSTRACT (cont from p.1) selection until, after nine rounds, the newly selected molecules functioned nearly as well as the original RNA molecules did. An additional five rounds of selection did not result in further improvement. RNAs present at various stages of selection were randomly chosen and sequenced. Many sequence variants remained at the end of the selection, suggesting that there are multiple solutions to the problem of constructing a tobramycin beacon aptamer. The sequence variants were analyzed for clues about structural requirements. Analysis of the sequences variants provided evidence that supports a hypothetical secondary structure. Evidence for additional interactions begins to outline the tertiary structure of the molecule. In addition to increasing our understanding of RNA structure/function relationships, the results of these experiments will aide in the design of future *in vitro* selection strategies.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Aptamer, beacon, evolution, *in vitro*, RNA, tobramycin	**46**
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

Abstract

In vitro selection exploits the principle of natural selection to isolate nucleic acid molecules (RNA or DNA) that can perform a desired function. The method begins with a large collection of molecules, each with a different sequence (variation). Molecules that perform a particular function are then purified from the nucleic acid pool (selection). The selected molecules are replicated to make a new "generation" that is subjected to another round of selection. For example, *in vitro* selection is commonly used to produce RNA "aptamers" that bind tightly and specifically to a given target molecule.

"Beacon aptamers" are modified aptamers that change conformation and emit light upon binding to their target. Thus, beacon aptamers can function as very sensitive and specific probes for the detection and quantification of a wide variety of targets. The large number of potential applications of such probes has generated great interest in their efficient production. An *in vitro* selection strategy for the production of RNA beacon aptamers was recently devised and used to produce probes that could detect the antibiotic tobramycin.

This Trident project used the new *in vitro* selection strategy to explore the structure and function of one of the tobramycin beacon aptamers. The question addressed was the following: What are the sequence/structure constraints on a functional tobramycin beacon aptamer? The sequence of the RNA was partially randomized in order to produce a large number of variants related to the original aptamer. Then, the selection for functional molecules was repeated. The function and the sequences of the RNA molecules were monitored during the course of the selection process. The function of the RNA pools improved after each round of selection until, after nine rounds, the newly selected molecules functioned nearly as well as the original RNA molecules did. An additional five rounds of selection did not result in further improvement.

RNAs present at various stages of selection were randomly chosen and sequenced. Many sequence variants remained at the end of the selection, suggesting that there are multiple solutions to the problem of constructing a tobramycin beacon aptamer. The sequence variants were analyzed for clues about structural requirements. Analysis of the sequences variants provided evidence that supports a hypothetical secondary structure. Evidence for additional interactions begins to outline the tertiary structure of the molecule. In addition to increasing our understanding of RNA structure/function relationships, the results of these experiments will aide in the design of future *in vitro* selection strategies.

Keywords: Aptamer, beacon, evolution, *in vitro*, RNA, tobramycin

Acknowledgments

Professor Shade for her work as the Director of the Trident Scholar Program, the Trident Scholar Committee for their assistance in improving the project and the Office of Naval Research and the Naval Academy for funding this project.

Preface

Although defined the first time they are used within the text, many of technical terms are also located in the glossary following the report.

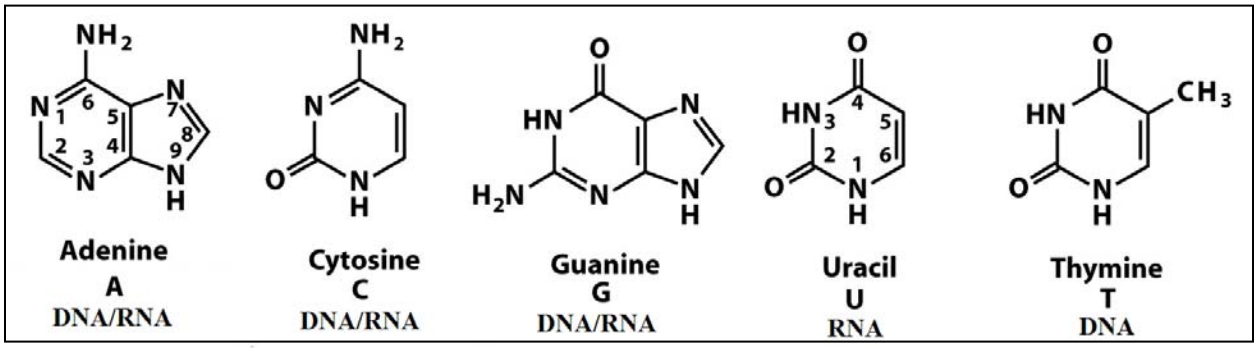
Table of Contents

Background

Structure and Function of RNA and DNA

Ribonucleic acid (RNA) and deoxyribonucleic acid (DNA) are linear polymers composed of the four monomers: adenosine 5'-monophosphate (A), cytidine 5'-monophosphate (C), guanosine 5'-monophosphate (G), and either uridine 5'-monophosphate (U, in RNA) or thymidine 5'-monophosphate (T, in DNA). These monomers are called nucleotides. Each nucleotide is composed of a sugar (ribose in RNA, deoxyribose in DNA), a phosphoryl group, and one of five nitrogen containing bases. The structures of the five bases are shown in Figure 1.

Figure 1: Bases Composing RNA and DNA[1]

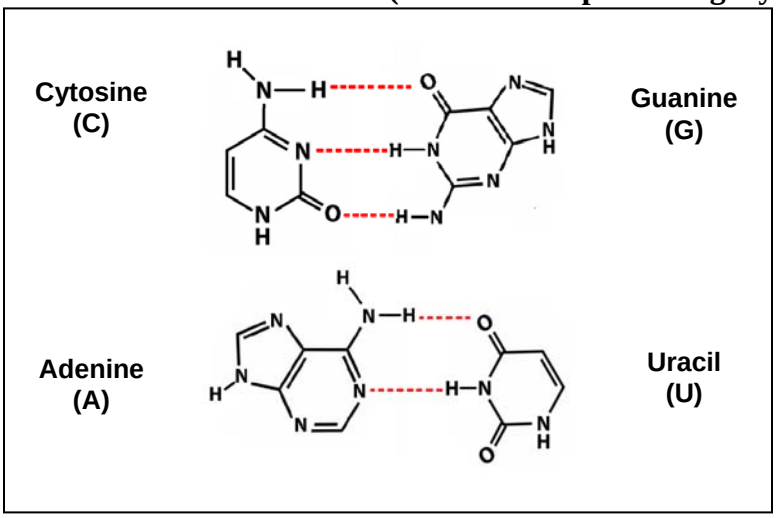


In both RNA and DNA, nucleotides are arranged like beads on a string, linked together through phosphodiester bonds. The structures of RNA and DNA are very similar, but there is one very important difference. RNA is usually single-stranded while DNA is composed of two separate strands held together through hydrogen bonds and other types of intermolecular forces. The DNA double-helix has a rigid rod-like structure, but the single strands of RNA can fold into complex shapes. In cells, DNA has a single function, storing genetic information. In contrast, RNA performs a wide range of functions. The nucleotide sequence of an RNA molecule determines its shape which, in turn, determines the function performed. The three-dimensional

[1] Morton, H. Robert; Moran, Laurence; Ochs, Raymond; Rawn, David; Scrimgeour, K. Gray. *Principles of Biochemistry.* Pearson Prentice Hall, Inc. 2006.

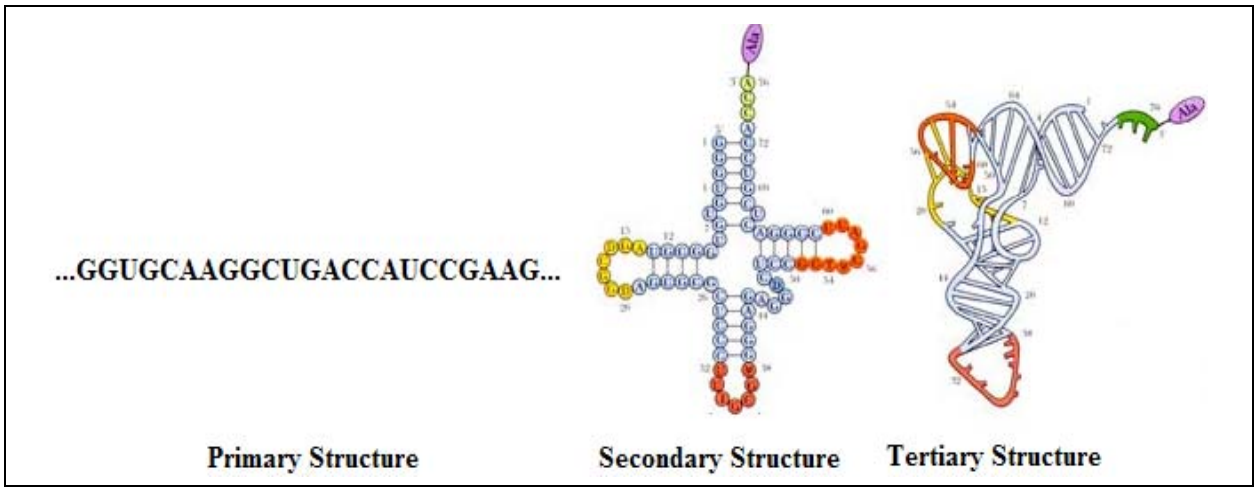
structure of an RNA molecule is stabilized by a variety of interactions between the monomers. Perhaps the most important stabilizing force is due to the phenomenon of "base pairing" in which hydrogen bonds form between two nucleotides. In RNA, U prefers to pair with A and C prefers to pair with G. C-G pairs are held together through three hydrogen bonds and U-A pairs are held together through two hydrogen bonds. Thus, C-G pairs are more stable than U-A pairs. Refer to Figure 2.

Figure 2: The Two Most Stable Base Pairs (Red Lines Representing Hydrogen Bonding)



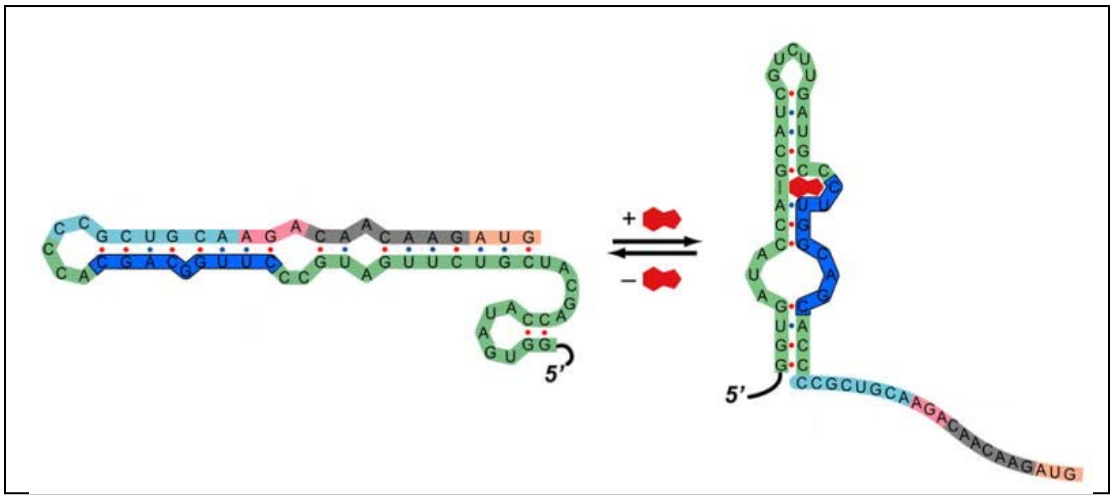
It is base pairing that keeps two complementary strands of DNA together, allowing it to take the famous double helix shape. Figure 3 gives an example of RNA folding. The RNA shown is a transfer RNA (tRNA) that is used in protein synthesis. The RNA's primary structure is the exact ordering of nucleotides; its secondary structure is the two dimensional form showing where the RNA is double stranded and where there are single stranded loops, but without showing the shape those loops will take. The tertiary structure is its overall three dimensional structure as it is in reality. Notice the lines between various nucleotides, representing base pairing in the double stranded regions.

Figure 3: RNA Primary, Secondary, and Tertiary Structure



While double-stranded DNA is rigid, RNA can undergo conformational changes. For example, many naturally occurring RNAs contain a "riboswitch" domain. A riboswitch binds to a specific molecule; this binding causes a change in conformation. The conformational change triggers downstream events that allow the cell to adapt to the presence of the ligand. Refer to Figure 4 for an example of a riboswitch assuming a different conformation once it binds its ligand (represented by the shape above the arrow).

Figure 4: Riboswitch Conformational Change[2]



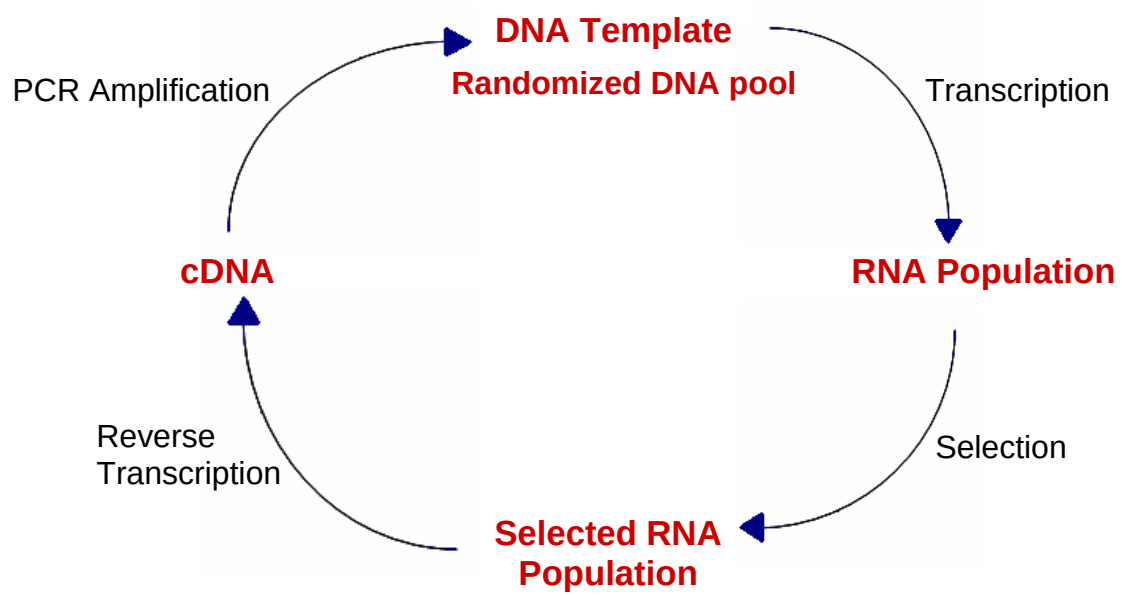
[2] Sean A. Lynch, Shawn K. Desai, Hari Krishna Sajja, and Justin P. Gallivan. A High Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into their Function. 2007. *Chemical Biology* 14(2): 173-184.

In Vitro Selection

Researchers have applied the concepts of natural selection to develop *in vitro* selection strategies to produce RNA molecules that perform a number of useful and interesting functions. One use of particular interest is as an "aptamer." An aptamer is any molecule that binds to a specific target molecule, or ligand. Through various means, the amount of bounded aptamer can be determined, and thus the amount of target molecule present measured.

Using the SELEX method, the production of RNA aptamers has become fairly routine. SELEX stands for **S**ystematic **E**volution of **L**igands by **Ex**ponential Enrichment (SELEX).[3] The method is outlined in Figure 5.

Figure 5: SELEX (*in vitro selection)* Mechanism Diagram



A ligand is bound to a solid phase to produce an affinity column. Numerous RNA molecules, each with a different sequence (randomized RNA pool), are passed through the column and the desired RNAs bind to the ligand. After washing the column with buffer to

[3] Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment. *Science*, 249:505-510.

remove unbound RNA, the bound RNAs are eluted (washed out) with excess ligand. The eluted RNA is converted into DNA by a process called "reverse transcription" (RT) and many copies of this DNA are made by "polymerase chain reaction" (PCR). The DNA is then converted back into RNA by "transcription" and this RNA is used as the starting point for another round of selection. SELEX is very similar to natural selection and the method is often referred to as *in vitro* evolution. A large variety of RNA sequences are present initially, and each time the sample is run through the column the poorly functioning sequences are removed. This process is repeated numerous times until a set of sequences that is better than all of the original molecules has been identified, not through logic or deductive reasoning but by simply letting nature decide which molecules are best fit for the task. Thus, SELEX follows the basic course of evolution by natural selection: variation, survival, differential reproductive rates, and repetition.

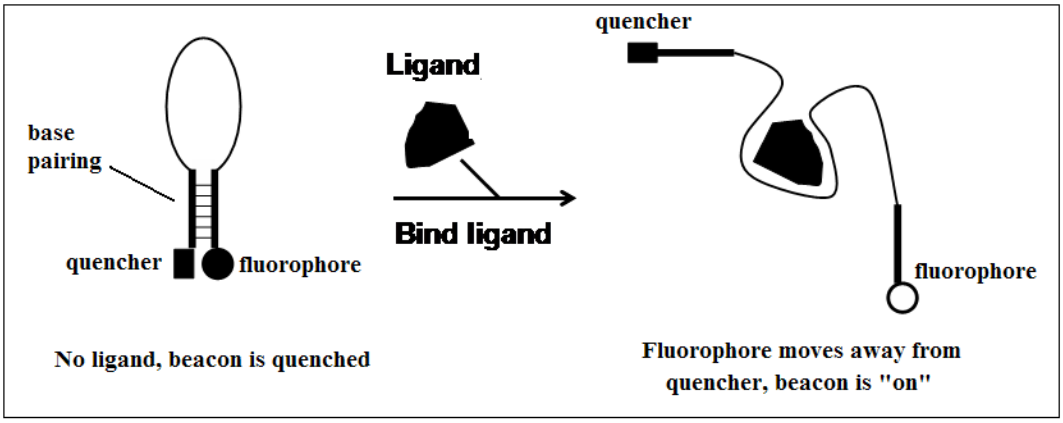
Beacon Aptamers and a New Selection Strategy

Beacon aptamers are a specific type of aptamer. They are RNA molecules that fluoresce, or emit light, once they have bound to a specific ligand.[4] In the unbound state, a beacon aptamer adopts a conformation in which a fluorophore and a quencher are next to each other (Figure 6). This conformation is stabilized through base pairing between the two ends of the RNA. A fluorophore is a molecule that emits light and a quencher is a molecule that absorbs the emitted light. When the fluorophore and quencher are next to each other, little to no fluorescence is emitted. When the ligand binds, the base pairing is disrupted, the aptamer undergoes a conformational change, and the fluorophore moves away from the quencher resulting in an increase in the fluorescence intensity. Thus, the fluorescence intensity is a

[4] Nutiu, R. and Li, Y. (2003). Structure-Switching Signaling Aptamers. *J Am Chem Soc* 125, 4771-4778.

measure of ligand concentration. Note that beacon aptamer function is very similar to that of naturally occurring riboswitches.

Figure 6: Mechanism of a Beacon Aptamer



In this project, beacon aptamers are being studied for their biochemical uses, but they could be used in a variety of applications. For example, they could be used in medicine to measure difficult to detect substances in the blood or could be used by the military to detect chemical or biological warfare agents present in an area. Detection of these agents is often difficult, expensive and most importantly slow, wasting precious time during which the agents can be infecting troops or civilians.

In the past, beacon aptamers were produced by modifying standard aptamers. Following selection of an aptamer, a fluorophore and quencher were attached to sites close to each other in the unbound conformation and the RNA was tested for its ability to signal the presence of the ligand through an increase in fluorescence intensity. This "rational design" approach has met with only limited success[5], most likely because most standard aptamers do not undergo the appropriate conformational change upon ligand binding.

[5] Yamamoto, R., Baba, T., and Kumar, P.K. 2000. Molecular beacon aptamer fluoresces in the presence of Tat protein of HIV-1. *Genes Cells* 5(5): 389-396.
Jhaveri, S., Rajendran, M., and Ellington, A.D. 2000. In vitro selection of signaling aptamers. *Nature biotechnology* 18(12): 1293-1297.

In order to more efficiently produce functional beacon aptamers, Dr. Morse devised a new selection strategy.[6] The new method (outlined in Figure 7) was designed to select RNA molecules with the two properties required of a beacon aptamer: ligand binding and the ability to undergo the appropriate conformational change. A similar method has been developed by Nutiu and Li.[7]

The new method is the reverse of the standard SELEX protocol in that the randomized RNA pool, rather than the ligand, is immobilized. RNA is bound to magnetic beads via base pairing with an oligonucleotide (small, single-stranded DNA molecule) that is tightly bound to the surface of the beads. The beads are washed to remove unbound RNA and then exposed to a solution containing the ligand of interest. RNAs that are released from the beads in the presence of the ligand are candidate beacon aptamers. Repeated rounds of selection further enrich the RNA pool for molecules that are released most efficiently. Note that release requires binding of the ligand to an RNA and disruption of the base pairing that holds the RNA on the beads. Such RNA molecules can be converted into beacon aptamers that can detect the ligand through an increase in fluorescence intensity. The details of each step of the procedure are described below. Dr. Morse has used this strategy to select RNA molecules that can detect the antibiotic tobramycin.

Hamaguchi, N., Ellington, A., and Stanton, M. (2001). Aptamer beacons for the direct detection of proteins. *Anal Biochem 294*, 126-131.

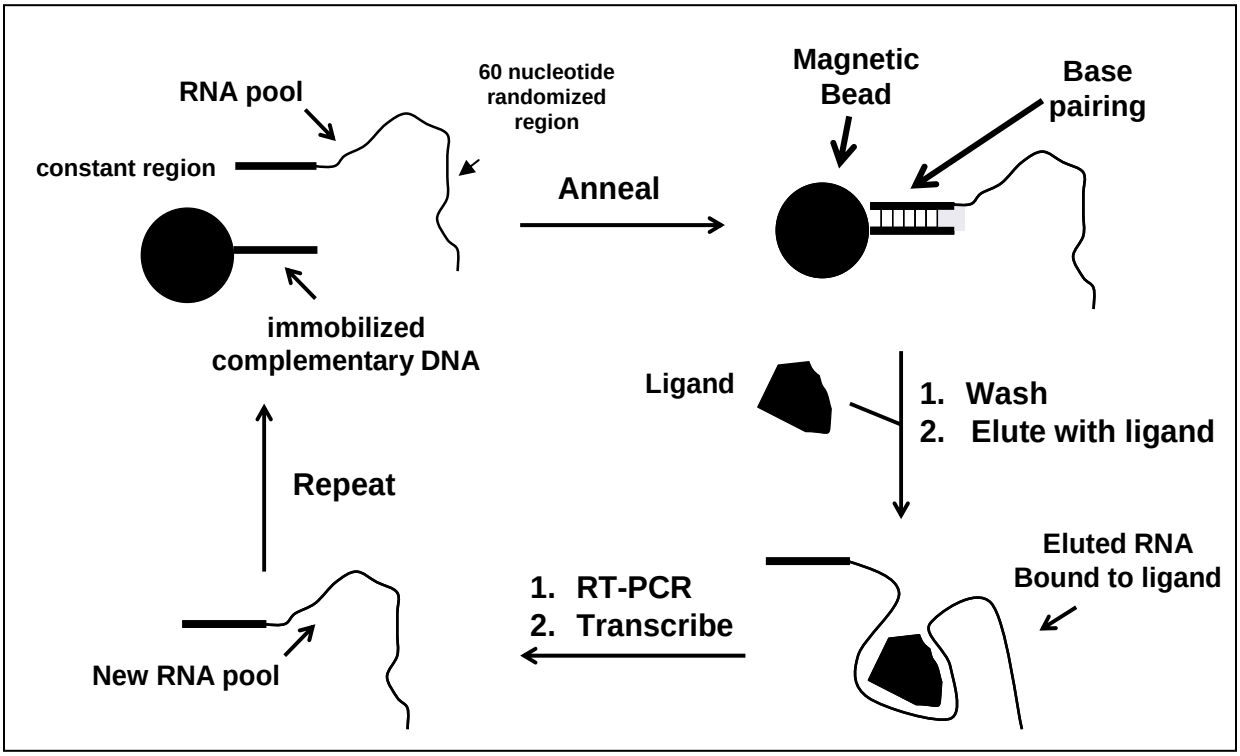
Frauendorf, C. and Jaschke, A. 2001. Detection of small organic analytes by fluorescing molecular switches. *Bioorganic & medicinal chemistry* 9(10): 2521-2524.

Li, J.J., Fang, X., and Tan, W. 2002. Molecular aptamer beacons for real-time protein recognition. *Biochem Biophys Res Commun* 292(1): 31-40.

[6] Morse, Daniel P. 2007. Direct Selection of RNA Beacon Aptamers. *Biochemical and Biophysical Research Communications 359*: 94-101.

[7] Nutiu, R. and Li, Y. (2003). Structure-Switching Signaling Aptamers. *J Am Chem Soc* 125, 4771-4778.

Figure 7: New Method for Beacon Aptamer Selection



Structure/function analysis through sequence comparisons

Structure/function relationships within an RNA can be studied by comparing the sequences of a collection of functional variants. A specific nucleotide in an RNA can play a variety of functional roles. It may be important for stabilizing a particular conformation through base-pairing or other types of interactions. It may interact with a ligand, or it may simply act as a spacer to allow various interactions to occur with the correct geometry.

Nucleotides that are base-paired can be identified by "covariation" analysis if the pairing is important for function[8]. This type of analysis relies on the fact that G-C pairs can sometimes be replaced by A-U pairs (or vice-versa) without significantly altering the RNA function. If a functionally important base-pair is disrupted by a nucleotide change in a molecule, then such molecules will compete poorly during *in vitro* (or natural) selection. A second nucleotide change

[8] Gultyaev, AP; Franch, T; Gerdes, K. 2000. Coupled Nucleotide Covariations Reveal Dynamic RNA Interaction Patterns. *RNA*. 6(11), p. 1483-91.

that restores base-pairing should also restore function so such molecules should survive selection. The presence of "covariation" (two nucleotide changes that maintain the potential for base-pairing) within a collection of functional sequence variants provides strong evidence for the existence of specific base-pairs in the structure.

Sometimes, the stability of RNA structure is so fine-tuned that a G-C pair cannot functionally replace an A-U pair. Beacon aptamers likely include regions of this type since they are poised to convert between two alternative conformations upon binding to a ligand. Regions of this type will likely remain invariant following selection. Invariant nucleotides in single-stranded regions of an RNA may make very specific contacts with other regions of the RNA or with other molecules. Nucleotides that vary in an apparently random fashion likely serve as spacers and do not participate in functionally important interactions.

Detailed Description of *In Vitro* Selection Procedure

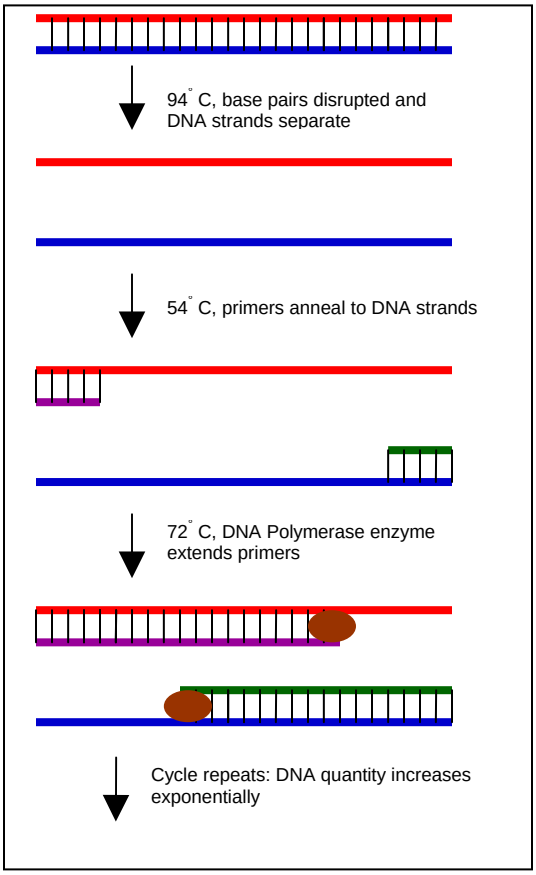
The *in vitro* selection process begins with a large collection of DNA molecules, each with a different sequence. For example, the following piece of DNA was synthesized commercially and used for the selection of tobramycin beacon aptamers.

GGAATGGATCCACATCTACGAN$_{60}$TTCACTGCAGACTTGACGAAAAGC

It is composed of a sixty nucleotide randomized region (N_{60}) flanked by two constant regions. The six nucleotides shown in bold are complementary to an immobilized DNA molecule (see below). Each position in the randomized region has an equal probability of being either A, G, C, or T. Thus, there are an extremely large number of possible sequences from which to select. The constant regions never changes during a selection experiment.

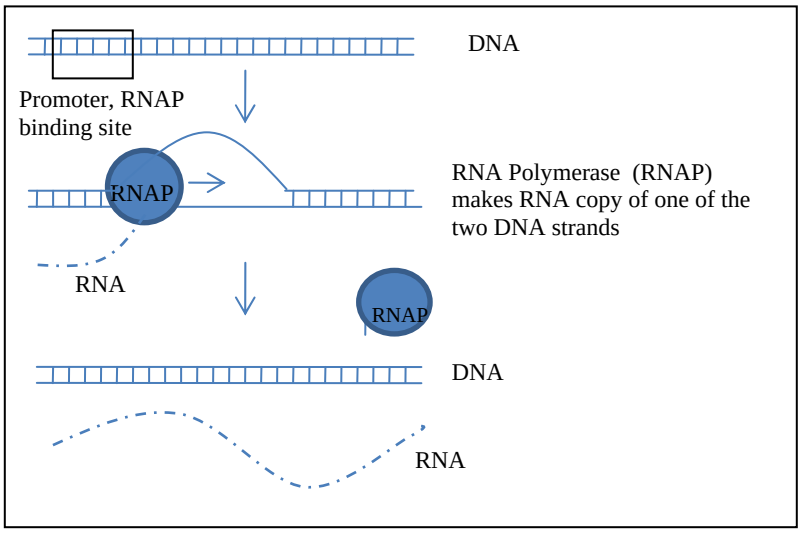
The single-stranded DNA is converted to double stranded DNA and amplified using the polymerase chain reaction, or PCR. Figure 8 shows how PCR exponentially replicates a specific DNA sequence. The constant regions provide primer binding sites for this process.

Figure 8: Polymerase Chain Reaction (PCR)



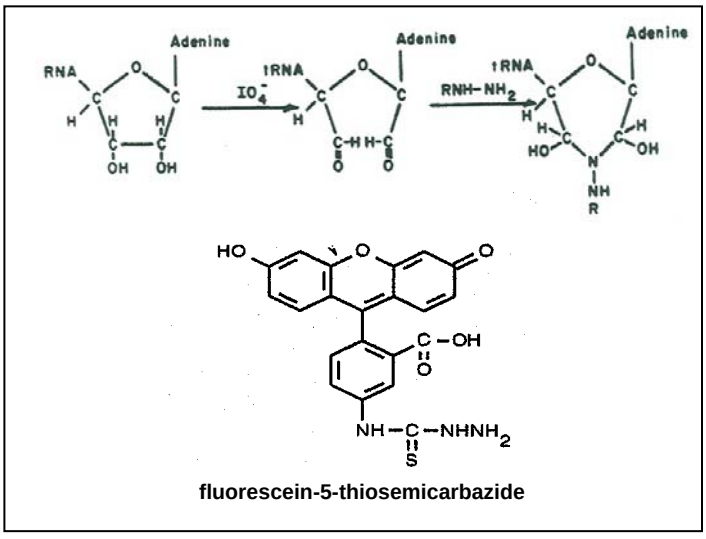
The DNA is then converted to single stranded RNA through the process of transcription, in which the enzyme RNA polymerase copies one of the strands of DNA. The transcription process is outlined in Figure 9. A promoter sequence, added during PCR, acts as a starting point for the RNA polymerase.

Figure 9: Transcription Mechanism



The RNA is labeled with fluorescein in order to quantify it at the end of each round of selection. The labeling procedure is shown in Figure 10. One end of the RNA is oxidized with periodate to make a dialdehyde. This structure then reacts with fluorescein thiosemicarbazide, as seen in the second reaction in Figure 10.

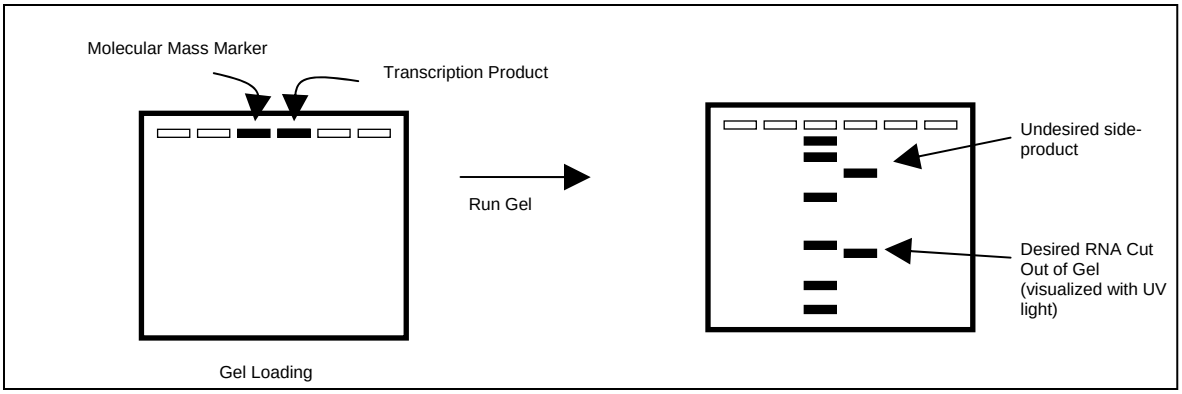
Figure 10: Oxidation and Labeling Reactions of RNA, and Structure of Fluorescein



The fluorescent RNA is then purified, including undergoing gel purification in which the product is run by electrophoresis on a polyacrylamide gel. Different molecules, depending on

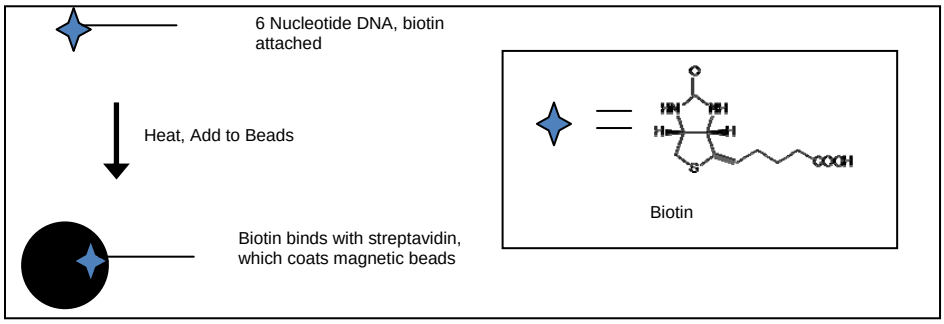
their size and makeup, run at different speeds on the polyacrylamide gel, which allows for the desired full length fluorescent RNA to be visualized with UV light and cut out of the gel. This ensures that in future steps reaction components from previous reactions do not interfere. This process is illustrated in Figure 11.

Figure 11: Acrylamide Gel Purification of RNA



A small DNA molecule is attached to magnetic beads, as shown in Figure 12. This DNA is complementary to the six nucleotides at the end of each RNA in the randomized RNA pool described above. Biotin is attached to one end of the DNA strands. The biotin forms a very strong bond to streptavidin, which coats the magnetic beads. What results are DNA strands that are essentially immobilized on the magnetic beads.

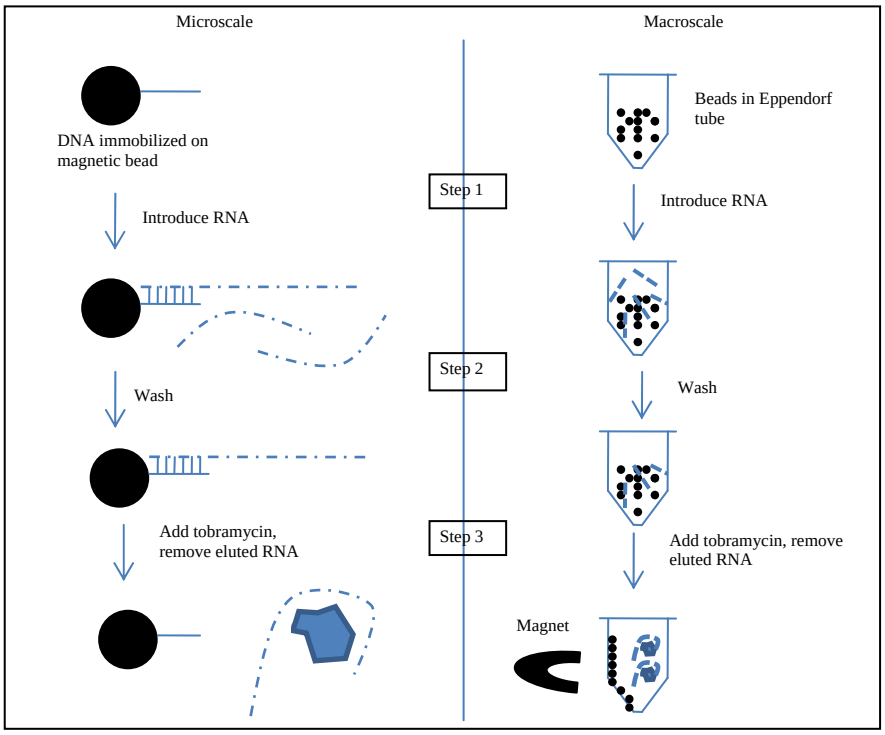
Figure 12: Attaching DNA to the Beads Using Streptavidin and Biotin



The randomized RNA pool is bound to the magnetic beads through base pairing between the complementary immobilized DNA and one end of the RNA (Figure 13, Step 1). This is the same type of interaction that stabilizes the unbound conformation of a beacon aptamer (compare Figure 13 to Figure 5). After washing to remove unbound RNA (Figure 13, Step 2), bound RNA molecules are eluted with the desired ligand (Figure 13, Step 3). Elution requires that an RNA not only bind to the ligand but also that binding disrupts the RNA-DNA interaction. This is analogous to the conformational change required of a beacon aptamer (compare Figure 13 to Figure 6). In both cases there is base pairing that is being disrupted upon ligand binding. The bound and unbound RNAs are separated by capturing the magnetic beads with a magnet and removing the solution containing unbound RNA (Figure 13, Step 3). RNA that remains bound to the beads is removed by heating the beads to 65^{o} C for ten minutes. The percent elution of the RNA is then determined by measuring the fluorescence intensities of the eluted RNA and of the RNA that remained bound to the beads, and using the equation:

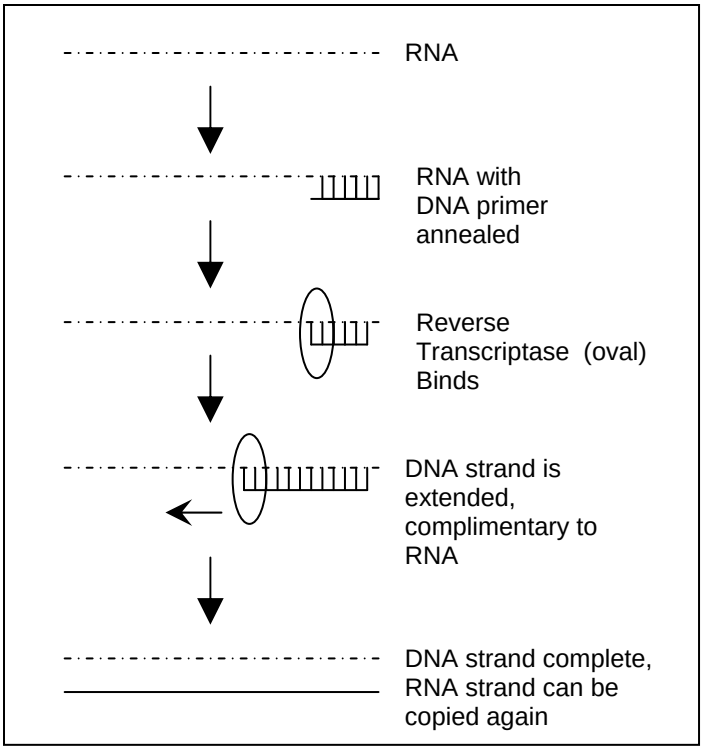
$X\% = \frac{F_{elute}}{F_{elute} + F_{bound}}$	F_{elute} = fluorescence of eluted RNA F_{bound} = fluorescence of bead-bound RNA X% = percent RNA eluted from the beads

Figure 13: Binding RNA to Beads and Elution with Tobramycin



As in the standard SELEX approach, the eluted RNA is converted to DNA through the process of reverse transcription (RT), amplified with PCR, and converted back to RNA for the next round of selection. Refer to Figure 14 for the RT process, Figure 8 for the PCR process, and Figure 9 for the transcription process. Note that the single stranded DNA at the end of the RT process is used at the beginning of the PCR process.

Figure 14: Reverse Transcription (Convert RNA to Single Stranded DNA)

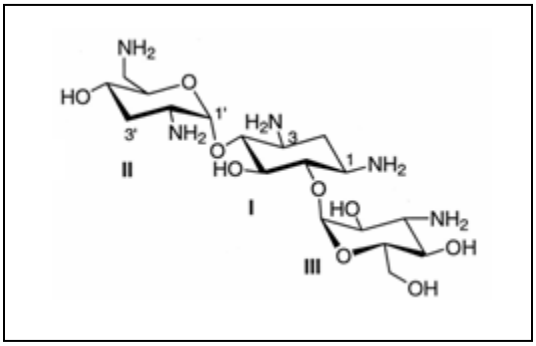


The identities of the eluted RNAs can be determined by converting them to DNA followed by cloning and DNA sequencing. First, RT-PCR produces many DNA copies of the RNA. Then cloning is used; cloning is a method that uses bacteria to isolate individual DNA sequences, allowing them to be purified and sequenced commercially. Each of the purified DNA molecules encodes a single RNA. Therefore, individual RNAs from the eluted pool can be synthesized and tested for function by measuring their efficiency of elution from the magnetic beads. Alternatively, the function of individual RNAs can be tested in the beacon aptamer format (see Figure 6). It requires several additional steps to convert selected RNAs into beacon aptamers so, in this project, function was tested by measuring elution efficiency.

Previous Work

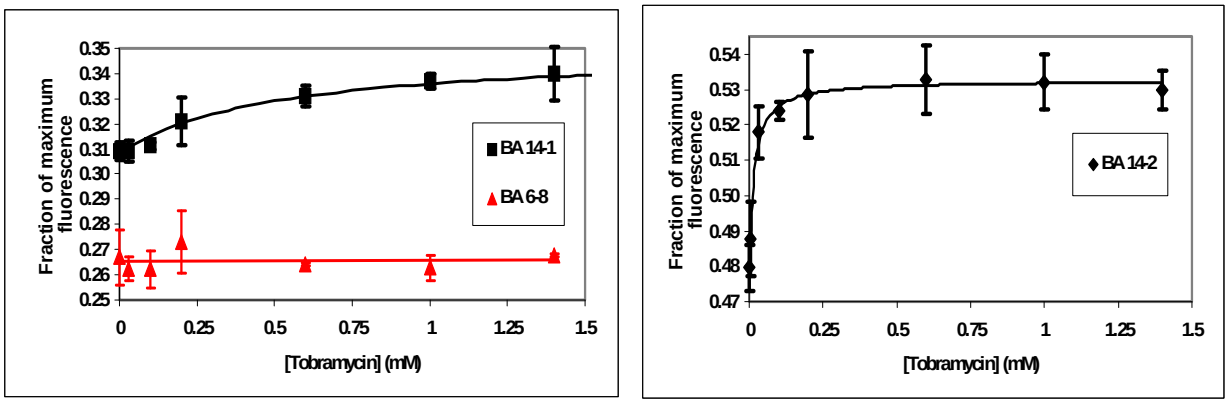
To test the new method of *in vitro* selection, the antibiotic tobramycin was selected as the target molecule (Figure 15). Tobramycin was chosen to be used in the initial proof-of-principle experiment for two reasons. Tobramycin has been found to bind extremely tightly to other aptamers, and there are also several molecules with similar structures that can be used in counter-selection, which ensures the beacon aptamer's specificity.

Figure 15: Structure of Tobramycin



After fourteen rounds of selection, only two unrelated RNA sequences remained, that were named 14-1 and 14-2. When these RNAs were converted to beacon aptamers, they were both able to detect tobramycin but their performance was suboptimal. An optimal beacon aptamer would be able to detect sub-micromolar concentrations of ligand. The 14-1 and 14-2 beacon aptamers could detect no less than 200 micromolar and 30 micromolar, respectively, and the fluorescence intensity of both aptamers increased only 10% at saturating concentrations of tobramycin (Figure 16, BA 14-1 and BA 14-2). Importantly, an RNA that did not survive the selection process did not function as a tobramycin beacon aptamer (Figure 16, BA 6-8). These results show that 14-1 and 14-2 were selected for their ability to detect tobramycin.

Figure 16: 14-1, 14-2 and 6-8 in Beacon Aptamer Format



Project Work

Aims

The question posed in this Trident Scholar Project was, "What are the sequence/structure constraints on a functional tobramycin beacon aptamer?" This question was addressed by generating RNAs that were closely related to 14-2 RNA and determining which variants work at least as well as the original. 14-2 RNA was chosen over 14-1 RNA for further study because it could detect a lower concentration of tobramycin. It is important to understand that variants related to 14-2 RNA (additional family members) were not isolated in the original selection because they were almost certainly absent from the initial randomized RNA pool. With a 60-nucleotide randomized region there are a total of $4^{60} = 10^{36}$ possible sequences; however, only 10^{14} different sequences were used for selection. Thus, it is very unlikely that any two of the 10^{14} molecules were closely related. It is also unlikely that 14-2 RNA was the best tobramycin beacon aptamer in its family since it was a randomly chosen member of the family. Therefore, a secondary goal of the project was to find a more efficient tobramycin beacon aptamer. In addition, it was hoped that the selection procedure would be improved along the way.

Generating 14-2 Family Members

The sequence of 14-2 RNA was partially randomized to create a pool of closely related family members from which functional variants would be isolated by repeating the selection procedure. This approach has been used successfully by others to improve the function of selected RNA molecules. For example, Bartel and Szostak used this type of "*in vitro* evolution" to greatly improve the function of a catalytic RNA molecule.[9]

The original selection experiment began with a synthetic DNA oligonucleotide with a 60 nucleotide region that was completely randomized flanked by two constant regions. To begin the new selection, a similar oligonucleotide was synthesized that was identical to the 14-2 sequence except that the central 60 nucleotides were 21% randomized. This means that each of the 60 positions had a 79% chance of being the original nucleotide present in 14-2, and each of the remaining three nucleotides had a 7% chance of replacing it. The new RNA pool produced from this oligonucleotide was named JV2. In order to confirm the partial randomization, the sequences of seventeen RNAs from the new pool were determined. (Twenty molecules were sequenced but three of the sequencing reactions failed.) In Figure 17 the randomized region (central sixty nucleotides) of the seventeen sequences is compared to the sequence of 14-2. The 14-2 sequence is shown at the top of the alignment (majority). The highlighted nucleotides are those that differ from the 14-2 sequence. The table in Figure 18 compares the expected frequency of each nucleotide to the frequencies observed in the seventeen sequenced molecules. There was on average eleven changes in each sixty nucleotide region when compared to the 14-2 sequence. This is very close to the average number of changes expected (13). However, the degree of randomization did not completely match expectations. For example, instead of the

[9] Bartel, D.P. and Szostak, J.W. (1993) Isolation of New Ribozymes from a Large Pool of Random Sequences. *Science*, New Series 261, 1141-1418.

expected 7%, T replaced G less than one percent of the time and G replaced A eleven percent of the time. However, after taking into account the fact that only seventeen molecules were sampled from the large population, it was concluded that the randomization was satisfactory and the experiment was continued.

Figure 17: Round 0

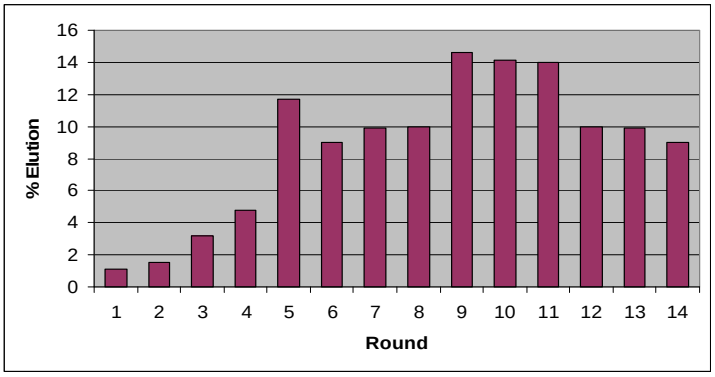
Figure 18: Desired and Actual Results of Partial Randomization

		Nucleotide in Original (14-2)			
		A	**C**	**G**	**T**
% Nucleotide Ordered Commercially	**A**	**79%**	7%	7%	7%
	C	7%	**79%**	7%	7%
	G	7%	7%	**79%**	7%
	T	7%	7%	7%	**79%**
% Nucleotide in JV2-0 Pool	**A**	**76.50%**	8.00%	4.80%	3.60%
	C	5.00%	**83.00%**	6.70%	3.30%
	G	11.50%	2.00%	**88.30%**	7.10%
	T	6.90%	7.00%	0.20%	**85.90%**

Selection Progress

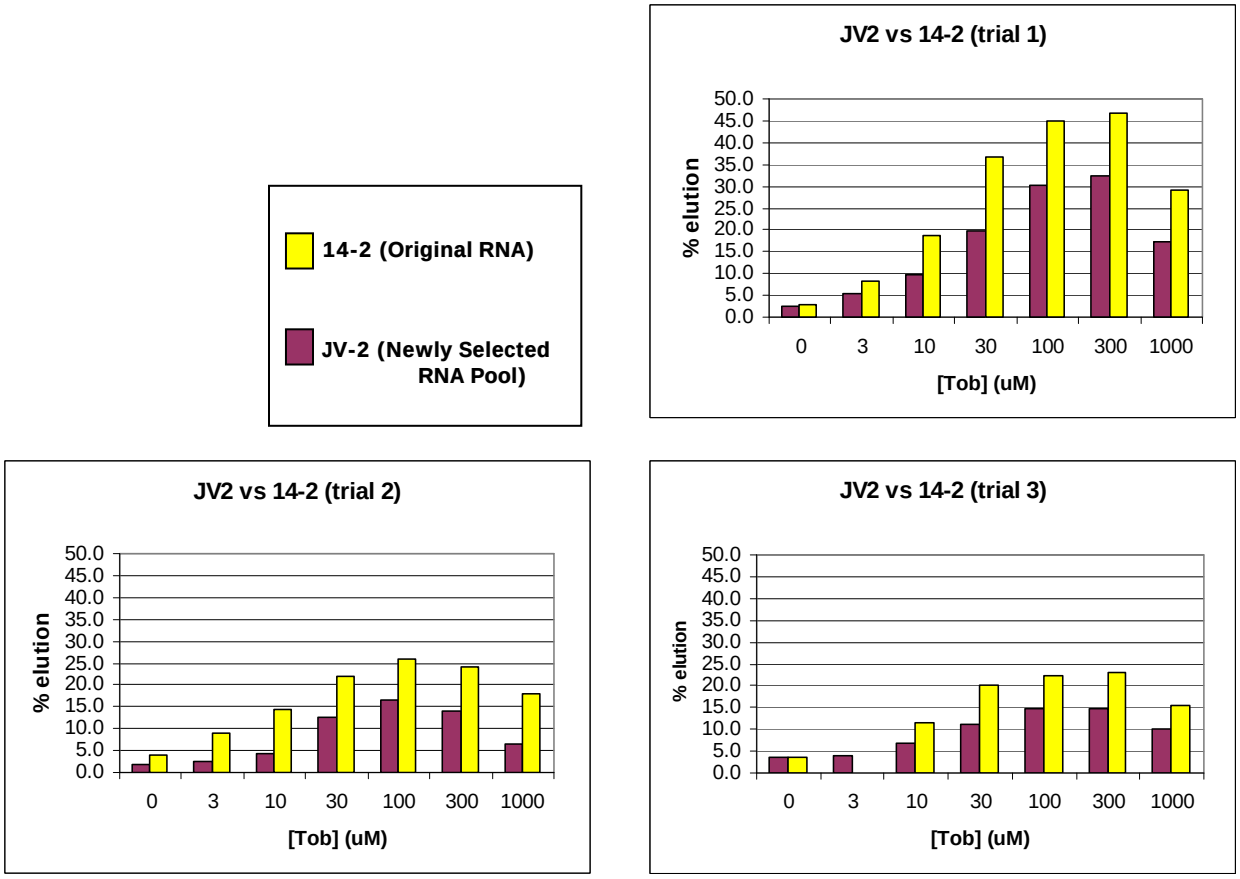
Fourteen rounds of selection were carried out, starting with the partially randomized RNA pool. In each round the RNA was eluted with 20 µM tobramycin which is the lowest tobramycin concentration used in the original selection experiment. After each round, the percent of bound RNA that eluted from the magnetic beads was calculated. The results are shown in Figure 19. Very little RNA eluted in the first round, suggesting that only a small fraction of the partially randomized pool would be able to perform the desired function. As expected, the elution efficiency appeared to increase as the selection progressed. It cannot be determined from this data if the observed changes in elution efficiency are statistically significant because the experiment could only be carried out once, due to the nature of the work.

Figure 19: Progress of Selection of Partially Randomized RNA Pool (JV2)



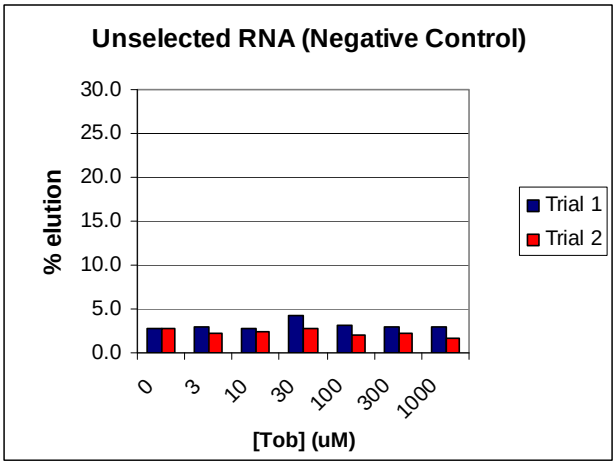
In order to more carefully assess the function of the newly selected RNA pool, the round 14 pool was studied in more detail. The elution efficiency as a function of tobramycin concentration was measured and compared to 14-2 RNA. The results are shown in Figure 20.

Figure 20: Elution Efficiency of Round 14 RNA Pool Compared to 14-2 RNA



As a negative control, the elution efficiency of an unselected RNA was also measured, as seen in Figure 21.

Figure 21: Elution Percentage of Unselected RNA



In three trials the newly selected RNA pool performed similar to but not as well as 14-2 RNA. Importantly, in two trials the unselected RNA did not respond to any concentration of tobramycin. Overall these results suggest that the new RNA pool does not contain RNAs with significantly improved function over 14-2 RNA. However, the results also demonstrate that the RNAs were, in fact, selected for their ability to respond to tobramycin by undergoing a conformational change.

The results also confirmed the previous observation that at high concentrations of tobramycin the elution efficiency decreases. This is most likely because tobramycin is weakly binding non-specifically to RNA and stabilizing the pairing between the RNA and the immobilized DNA oligonucleotide. It has been shown that aminoglycosides such as tobramycin can stabilize RNA structures.[10] Thus it seems that there are two types of binding sites for tobramycin. At low concentration it binds specifically to one or more high affinity sites and causes a conformational change. However, at high concentrations, tobramycin binds to low affinity sites and stabilizes the RNA-DNA duplex, making it more difficult for the RNA to elute.

It was originally hoped that the new selection would yield RNA molecules that would function significantly better than 14-2 RNA as tobramycin beacon aptamers. Although it appears that the final pool of RNA performs slightly worse than 14-2 RNA, it is still possible that individual variants within that pool function better. There are several possible explanations for the apparent lack of improvement. It may be that no family member works as well as 14-2 RNA, so partially randomizing it would never have a positive effect. If this was the case, 14-2 would be the dominant RNA sequence present in the final RNA pool. It is more likely that the final pool contains a variety of closely related family members, all with comparable efficiencies. On the

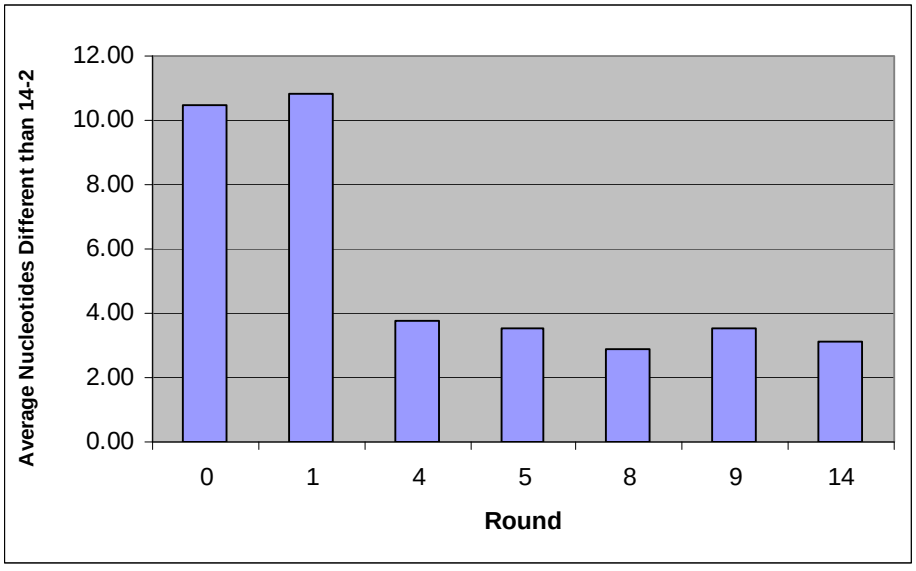
[10] M. Kaul, D. Pilch. Thermodynamics of Aminoglycoside-rRNA Recognition: The binding of neomycin-class aminoglycosides to the A site of 16S rRNA, Biochemistry 41 (2002) 7695-7706.

other hand, the degree of partial randomization may not have been large enough to allow improved variants to effectively compete with the original 14-2 sequence. However, this is most likely not the case as the original RNA pool was well randomized and 14-2 RNA was not present among the seventeen sequenced molecules. Finally, about twenty nucleotides at each end of the RNA were not allowed to change in order to provide priming sites for reverse transcription and PCR. This could have prevented other parts of the molecule from changing if they interact with the constant regions, such as through base pairing. This is probably the most significant factor that could have prevented the experiment from isolating improved variants.

Structure/Function Analysis

In order to follow the progress of the selection experiment and to reveal the nature of the final RNA pool, the sequences of twenty RNAs present after rounds 1, 4, 5, 8, 9, and 14 were determined. It must be stressed that twenty sequences out of possibly thousands will yield only a rough approximation of the composition of each RNA pool. Figure 22 shows how the complexity of the RNA population changed during the selection.

Figure 22: JV2 Nucleotide Changes from 14-2

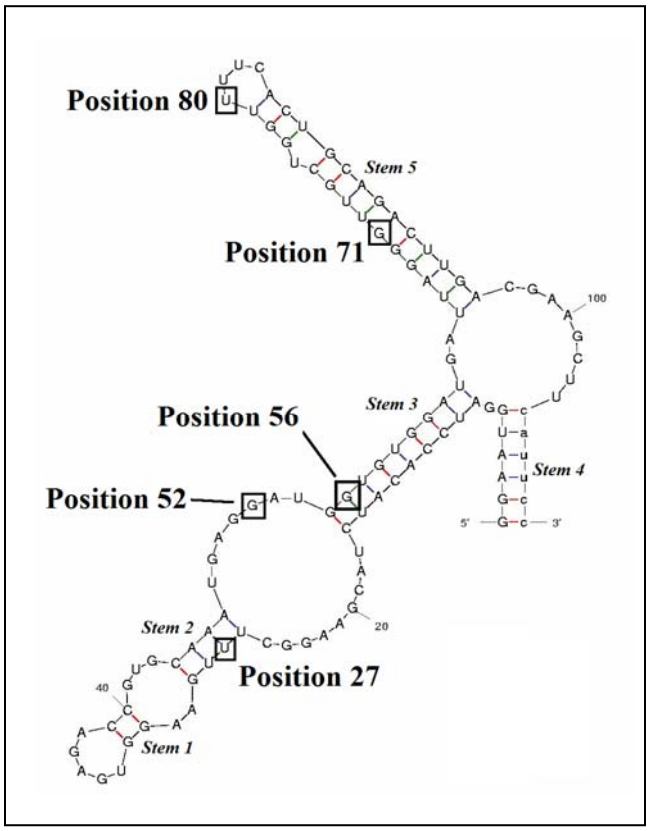


Rounds 0 and 1 were chosen to be sequenced to learn about early selection stages, 4, 5 and 8 and 9 because of the jump in elution efficiency between these rounds, and 14 to see a sample of the final products of the selection. After round 1, on average, about eleven of the sixty nucleotides in the randomized region were different than those found in 14-2 RNA. This is similar to the complexity of the original partially randomized pool. By round 4, the complexity was greatly decreased. An average of only four nucleotides were different than 14-2. Notice that the drop in complexity in round 4 correlates with increased elution efficiency in round 5 (see Figure 19). The complexity remained at about this level for the rest of the selection experiment. Significantly**,** the original 14-2 RNA sequence did not dominate the final RNA pool. (Only one of the twenty sequences was identical to 14-2.) Instead, as anticipated, a significant amount of variation remained. These results suggest that the vast majority of RNAs in the original pool functioned poorly and were quickly eliminated from the population. Although the average complexity and elution efficiency did not change significantly after round 4, it was still possible that the population continued to be enriched for individual variants with improved function.

The presence of sequence variants in the final round of selection provided the opportunity to examine structure/function relationships within this RNA family. Figure 23 shows a predicted secondary structure for 14-2 RNA. The last six nucleotides in the predicted structure shown in Figure 22 are those that were attached to the magnetic beads. This sequence is complementary to the first six nucleotides in the RNA. The boxed nucleotides are discussed below. The randomized region spans nucleotides 22 – 80. The two constant regions flank the randomized region. The structure was generated by a program called M-fold that uses thermodynamic data to predict secondary structure based on the primary sequence of an RNA. Note that M-fold predicts that the two ends of the RNA are base- paired as expected. Although such programs

work fairly well for small RNAs, the predictions represent only hypothetical structures that must be experimentally tested. Comparing the sequences of functional variants present in the selected

Figure 23: Predicted Structure of the JV2 Majority Sequence



RNA pool provides one means of testing the validity of the predicted secondary structure. The predicted structure is composed of five double-stranded "stems" (numbered 1 – 5 in Figure 23) connected by single-stranded "loops" and "bulges". Additional interactions, not shown in the structure, would be responsible for folding the RNA in three dimensions.

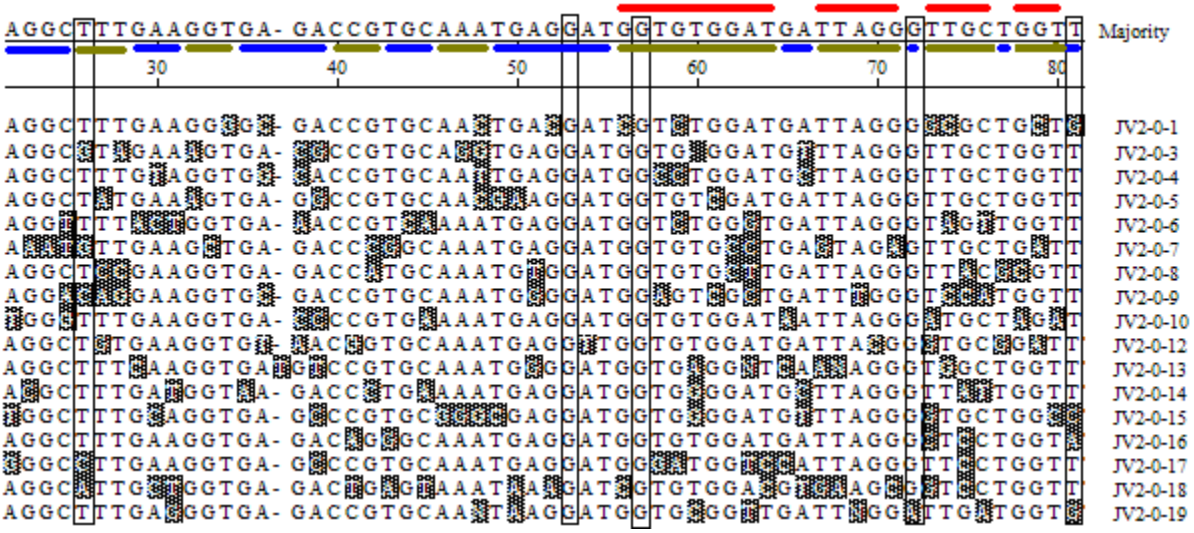
In Figure 24 the sequences of the twenty RNAs determined in each round are compared to the sequence of 14-2 (majority). Only the sixty nucleotide randomized region is shown. The constant regions that provided priming sites for reverse transcription and PCR are not included. The dramatic decrease in complexity starting in round 4 is apparent from these comparisons.

The alignments also show clear evidence for strong selection pressure in various regions of the RNA. There are three clear patterns: 1) regions where all twenty sequences in the final pool are identical to 14-2; 2) regions where a particular nucleotide effectively competed with the nucleotide present in 14-2; and 3) regions where the original sequence diversity was maintained.

Figure 24: JV2 Sequences

Positions 27, 52, 56, 71 and 80 are boxed in order to better see the changes from round to round. These positions are discussed later. Regions where the majority sequence (14-2) is underlined in blue are predicted to be single stranded, whereas yellow lines indicate predicted double stranded regions. The regions over-lined in red are complementary to the constant regions. Differences from 14-2 (labeled as the majority) are highlighted in black.

<u>Round 0</u>



Round 1

Round 4

Round 5

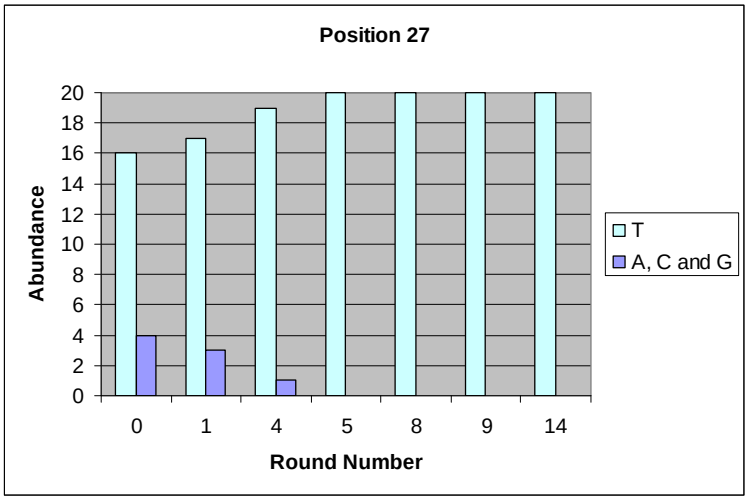
Round 8

Round 9

Round 14

At some nucleotide positions there was a clear progression from variation in round 0 to reversion back to the original nucleotide, most likely showing that only those sequences with that particular nucleotide at that particular position functioned well. Position 27 is an example of this (pattern 1); T is the majority nucleotide, and it is all that is left from round 5 on despite the diversity in the original RNA pool. Figure 25 shows how nucleotides other than T at position 27 disappear as the selection progressed.

Figure 25: JV2 Position 27 Sequence Summary

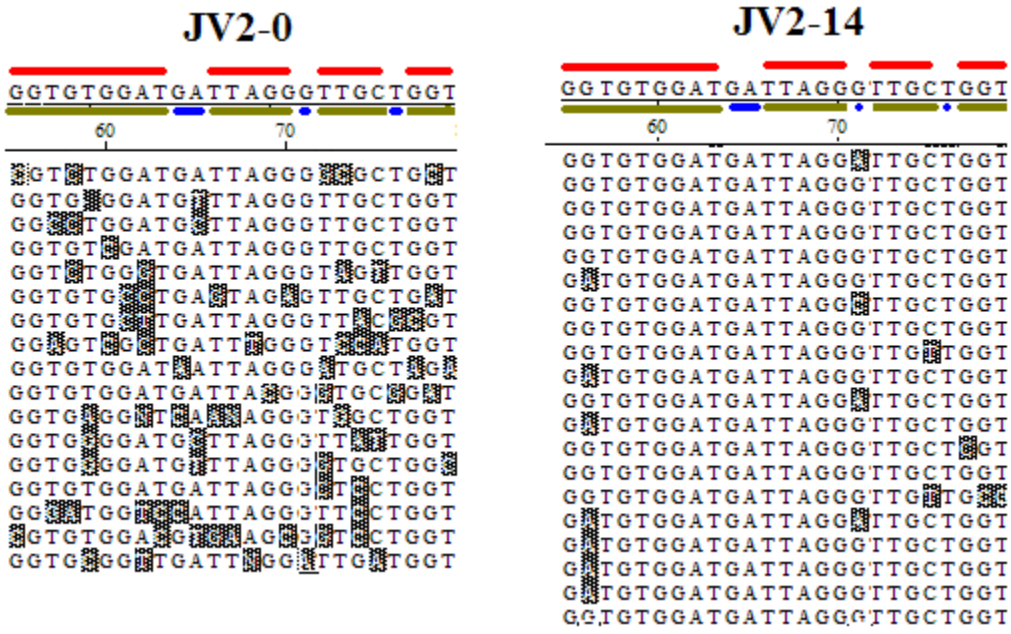


This progression suggests that changing the identity of the position 27 nucleotide results in a significant decrease in elution efficiency, causing very few of them to be selected each round.

In the final RNA pool, there is almost no variation in the sequence from positions 55 to 79. This could be due to the fact that this region is predicted to pair with the constant regions of the RNA (stems three and five in Figure 23). If this pairing actually occurs, and is functionally important, one would predict that positions 55 to 79 would not be allowed to change. This is exactly what is observed. Figure 26 compares this region in the original and final RNA pools.

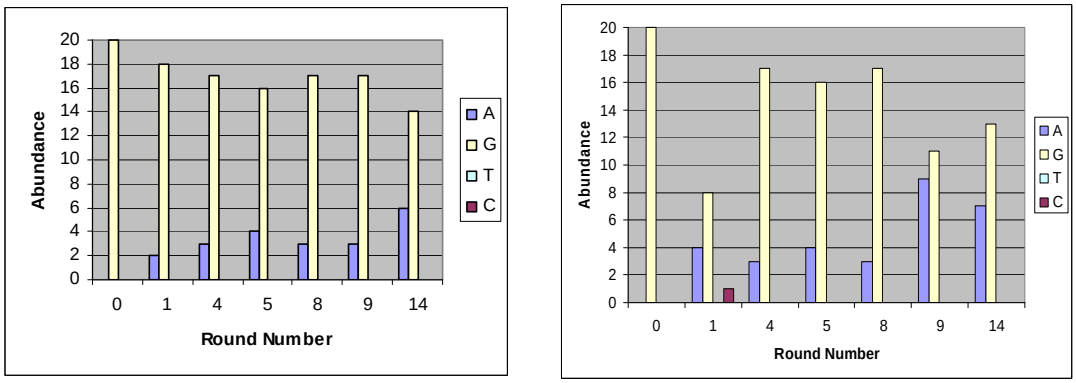
(As noted in Figure 24, blue and green lines indicate predicted single- and double-stranded regions, respectively. Red lines indicate nucleotides that are predicted to pair with the constant regions of the RNA.)

Figure 26: Evidence for the presence of stems 3 and 5



Positions 52 and 56, however, displayed pattern two behavior (selection for a specific nucleotide change). At both positions G was present in all twenty sequences in round 0. However, both positions had a steady increase in the abundance of A as the selection progressed. (Figure 27). This competition suggests that the competing nucleotide may allow for improved function, but this hypothesis remains to be tested.

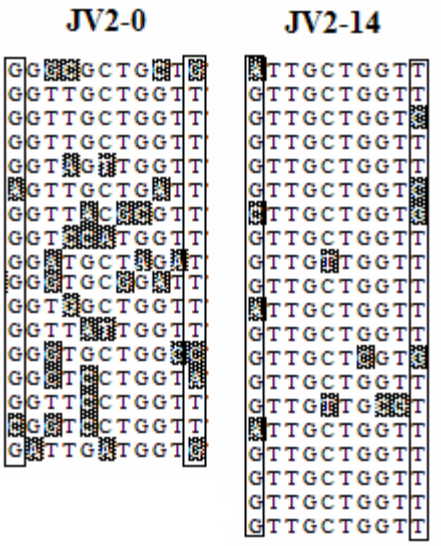
Figure 27: Selection for A at Positions 52 (left graph) and 56 (right graph)



Position 52 is predicted to be in the single stranded region (refer to Figure 23) and so may take part in binding tobramycin or stabilizing the tertiary structure of the molecule. Position 56 is predicted to be in a double stranded region. The G at position 56 is predicted to form a G-U pair within a nine base-pair stem. G-U pairs are commonly found in RNA, but they are not as stable as either an A-U or G-C pair. The other eight pairs in the predicted stem are either A-U or G-C. Changing position 56 from G to A would replace the G-U pair with a more stable A-U pair. This analysis predicts that stabilizing the stem may improve the function of the RNA. If there were covariation in this region that maintained the potential for formation of the nine base-pair stem, it would provide strong evidence for the existence of this stem. However, half of the stem is part of the constant region, so covariation could not arise during the selection. The invariance of positions 55, 57-63, coupled with the position 56 G to A, change provides less direct, but compelling, evidence for the presence of the stem.

Covariation analysis provides evidence for base-pairing between G-71 and U-80.

Figure 28: Evidence of Base Pairing Between Positions 71 and 80



Both nucleotides are predicted to be single stranded. G-71 is imbedded in a predicted stem but it is opposite an A. A G-A pair is highly unstable so G-71 is probably free to pair with some other nucleotide. The final RNA pool contains multiple variants at positions 71 and 80 but 19 of the 20 sequences maintain the potential for pairing between them. Whenever position 71 changes to an A, position 80 remains a T. (This occurs in 3 of the 20 sequences.) Whenever position 80 changes to a C, position 71 remains a G (2 sequences). Importantly, there is one sequence where both positions change. In this case, position 71 changes to a C and position 80 changes to a G. Only one of the 20 sequences in the final pool does not have complementary nucleotides at positions 71 and 80. (They are both G.) In the original RNA pool, only two of the six variants maintain the potential for base pairing.

Improvement of the selection procedure

Extensive work was done to determine the best parameters for the selection. PCR conditions were optimized and the appropriate time of elution with tobramycin was carefully evaluated. Neither of these parameters had been carefully studied previously. It was important to optimize these factors before moving forward with the selection to ensure that proper RNA diversity would be maintained and that the selection would progress as intended, filtering out only those sequences that do not perform the desired functions.

Perhaps the most significant change implemented in the selection protocol was the use of fluorescein-labeled RNA rather than radiolabeled RNA. This greatly improved the ease of handling and facilitated the measurement of elution efficiency. It was discovered that fluorescein-labeled RNA bound irreversibly to the surface of the magnetic beads that had been used previously. These beads had a hydrophobic surface that interacted strongly with fluorescein. Switching to beads with a hydrophilic surface solved this problem.

The results shown in Figure 20 suggest a way to improve the reproducibility of the functional analysis of selected RNAs. The magnetic beads were reused several times and then replaced with fresh beads. Beads that had been used at least one time gave reproducible results but each time fresh beads were used, the elution efficiency increased. The reasons for this are not clear but this observation suggests that all future experiments should be performed with beads that have been used at least one time.

Discussion and Future Work

In this Trident Scholar Project a selection was performed using a newly developed method of *in vitro* evolution. The starting point was a partially randomized RNA pool based on the sequence of a previously selected tobramycin beacon aptamer called 14-2. The goals of the project were to learn about the structure/function relationships that exist in this beacon aptamer and to improve the *in vitro* selection method. Both goals were achieved.

Analysis of the functional sequence variants isolated by the selection provided significant insight into the structure of 14-2 RNA. The data is largely consistent with a theoretical secondary structure generated by M-fold. In addition, covariation analysis provided strong evidence for a tertiary interaction that was not predicted by the folding program. Additional sequence data coupled with a more detailed statistical analysis of the sequence comparisons will likely yield further structural insights.

The structure/function implications drawn from the sequence comparisons require experimental verification. This can be readily achieved by synthesizing and testing the function of a variety of sequence variants. For example, analysis of positions 52 and 56 suggests that A to G changes at these positions will improve the function of the beacon aptamer. The covariation seen at positions 71 and 80 suggests that any variants at these positions that maintain the

potential for base pairing should function better than sequence changes that disrupt base pairing. The invariance of regions complementary to the constant regions suggests the presence of stems three and five in the predicted secondary structure. If this is the case, then sequence changes that disrupt these stems should interfere with function, and changes that restore the stems should restore function.

The apparent failure to isolate greatly improved beacon aptamers may have been due to the presence of rather long constant regions that appear to base-pair with a significant fraction of the partially randomized region. This paring may have prevented the isolation of potentially improved variants (as evidenced by the invariance of the regions complementary to the constant regions). This observation suggests a simple way to produce improved variants: decrease the length of the constant regions. This could be achieved by using smaller primers for RT-PCR. Alternatively, the constant regions could be completely eliminated by covalently attaching and subsequently removing priming sites. This would allow the entire 14-2 RNA sequence to be partially randomized and would greatly increase the chance of isolated improved variants.

Tobramycin is an antibiotic that inhibits bacterial protein synthesis by binding to ribosomal RNA. (Ribosomes are large complexes of RNA and protein that catalyze protein synthesis.) It is known that tobramycin and other aminoglycoside antibiotics can stabilize RNA structure which probably explains why high concentrations of tobramycin decrease the efficiency of elution of RNA from the magnetic beads (see Figure 20). This observation suggests that it will be possible to select much more sensitive beacon aptamers for ligands that do not stabilize RNA.

Bibliography

Bartel, D.P. and Szostak, J.W. (1993) Isolation of New Ribozymes from a Large Pool of Random Sequences. *Science*, New Series 261, 1141-1418.

Chen, Ying; Carlini, David; Baines, John; Parsch, John. 1999. RNA Secondary Structure and Compensatory Evolution. *Genes and Genetic Systems*, Vol 74, No. 6 p. 271-286.

Frauendorf, C. and Jaschke, A. 2001. Detection of small organic analytes by fluorescing molecular switches. *Bioorganic & medicinal chemistry* 9(10): 2521-2524.

Gultyaev, AP; Franch, T; Gerdes, K. 2000. Coupled Nucleotide Covariations Reveal Dynamic RNA Interaction Patterns. *RNA*. 6(11), p. 1483-91.

Hamaguchi, N., Ellington, A., and Stanton, M. 2001. Aptamer beacons for the direct detection of proteins. *Anal Biochem 294*, 126-131.

Jhaveri, S., Rajendran, M., and Ellington, A.D. 2000. In vitro selection of signaling aptamers. *Nature biotechnology* 18(12): 1293-1297.

Li, J.J., Fang, X., and Tan, W. 2002. Molecular aptamer beacons for real-time protein recognition. *Biochem Biophys Res Commun* 292(1): 31-40.

Morse, Daniel P. 2007. Direct Selection of RNA Beacon Aptamers. *Biochemical and Biophysical Research Communications 359*: 94-101.

"Polymerase Chain Reaction", Access Excellence Resource Center. Accessed 7 April 2008. <http://www.accessexcellence.org/RC/VL/GG/polymerase.html>

Nutiu, R. and Li, Y. (2003). Structure-Switching Signaling Aptamers. *J Am Chem Soc* 125, 4771-4778.

Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment. *Science*, 249:505-510.

Yamamoto, R., Baba, T., and Kumar, P.K. 2000. Molecular beacon aptamer fluoresces in the presence of Tat protein of HIV-1. *Genes Cells* 5(5): 389-396.

Zuker, Dr. Michael. "Mfold RNA Folding Program." Rensselaer Bioinformatics Web Server. Accessed 12 March 2008. <http://mfold.bioinfo.rpi.edu/>

Glossary of Terms

- Aptamer- an RNA molecule that binds tightly to a specific target molecule.
- Base pairing- the hydrogen bonding that forms between complementary nucleotides in DNA and RNA. Base pairing is the main force of stabilization for double stranded DNA. In DNA and RNA, guanine and cytosine are complementary and thus base pair; adenine base pairs with thymine in DNA and uracil in RNA. Numerous factors can disrupt this base pairing, including the RNA or DNA binding to another molecule or excessive temperatures.
- Beacon Aptamer- an RNA molecule that binds tightly onto a specific molecule and emits light upon doing so.
- Deoxyribonucleic acid (DNA)- molecules that are long linear polymers of the nucleotides adenosine 5'-monophosphate (A), cytidine 5'-monophosphate (C), guanosine 5'-monophosphate (G), and thymidine 5'-monophosphate (T). Often DNA is found double stranded, which means two polymers that are complementary (A and T form hydrogen bonds, G and C form hydrogen bonds) have matched up and formed a double helical structure. The exact sequence of the nucleotides determines the information the DNA contains and the function of any RNA molecules and proteins encoded by it.
- DNA Polymerase- an enzyme used to replicate DNA. It cannot start a new DNA strand from scratch but can extend existing nucleotide strands. The enzyme "reads" the existing DNA template and creates a complementary strand, identical to the partner strand of the template.
- Elute- the process of a chemical coming off of a column or magnetic beads in various selection procedures. In this project, elution is refers to RNA dissociating from the six

nucleotide DNA sequence that is immobilized on the magnetic beads when tobramycin is introduced.

- Elution time- Amount of time tobramycin is allowed to mix with RNA bound to magnetic beads before the supernatant is removed and the RNA that came off the beads quantified.
- Fluorophore- a molecule that will absorb and emit light at specific wavelengths.
- Ligand- a specific target molecule.
- Magnetic Beads- small beads coated with streptavidin that are used in the RNA elution work. Six nucleotide DNA strands with biotin attached are immobilized onto the beads through the incredibly strong binding between streptavidin and biotin. The end of the RNA then base pairs with the complementary DNA found on the beads. When tobramycin is introduced to the system, RNA molecules that bind tobramycin have the base pairing interaction disrupted dissociate from the bead. A magnet can then be used to pull the beads out of the way and remove the supernatant, which includes the RNA that eluted from the beads.
- Nucleotide- a chemical building block for DNA and RNA. Each nucleotide is made up of a sugar (deoxyribose in DNA and ribose in RNA), a phosphate group, and one of four nitrogen-containing bases. The bases adenine, guanine and cytosine (A, G, C) are common to DNA and RNA while the remaining base in DNA and RNA are thymine or uracil, respectively.
- Oligonucleotide- a short strand of nucleotides, either RNA or DNA, which is normally shorter than twenty bases
- Partial randomization- DNA oligonucleotides are commercially synthesized as follows. The first nucleotide in the chain is attached to a solid support. The second nucleotide is

added and is covalently attached to the first. Unreacted nucleotides are removed by washing with buffer. The process is repeated until DNA with the desired sequence and length is produced. A large collection of random sequences can be produced by adding a mixture of all four nucleotides at each step of synthesis. The degree of randomization can be controlled by altering the ratios at which the four nucleotides are mixed.

- Polymerase Chain Reaction (PCR)- a process used to amplify large amounts of DNA. DNA polymerase is used to extend short DNA primers into full length DNA strands that are copies of the original template the primer was annealed to, and then both the strands can be used as templates, thus initiating a chain reaction. Thermal cycling is often used; each cycle includes several temperatures to separate the double stranded DNA, allow the primers to anneal, and permit the enzyme to extend the DNA.
- Primer- a short strand of single stranded DNA used in DNA amplification. DNA polymerase cannot create DNA from scratch but needs a starting point, and the primer serves this purpose. Used extensively in PCR.
- Quencher- a molecule that absorbs light at a specific wavelength.
- Reverse Transcription (RT)- the process of converting RNA into double stranded DNA.
- Ribonucleic acid (RNA)- RNA, like DNA is a linear polymer of nucleotides. It differs from DNA as follows: it is usually single-stranded, thymine is replaced by uracil, and deoxyribose is replaced by ribose. The exact linear sequence of bases (adenine, guanine, cytosine, and uracil (A, G, C, U) determines the structure and function of RNA.
- Round of Selection- The process of starting with a pool of DNA, converting it into RNA, labeling it with fluorescein, binding it to the magnetic beads, eluting with tobramycin,

and finally converting back into DNA. Several rounds of selection make up one entire selection process.

- Selection Pressure- the parameters used during RNA elution that determine what amount of RNA elute from the beads. Elution time and tobramycin concentration are the main factors. For example, lowering the elution time will allow selection of only those RNA molecules that bind the tobramycin and dissociate from the beads quickly, which are ideally the best at performing those desired functions because they did it fastest. However, increasing the selection pressure too quickly could cause successful molecules from being selected; by chance the poorer functioning molecules may have eluted while the best sequences were left on the beads.
- Selection Process- the process of starting with a pool of randomized RNA and performing several rounds of selection in order to find one or more RNA sequences that are able to perform the desired function.
- Supernatant- normally refers to any liquid above solids or precipitates in a solution. In this work it refers to the solution containing the tobramycin and RNA that eluted from the magnetic beads
- Tobramycin- an antibiotic that is used to treat different infections. It was selected as the ligand in this proof-of-principle experiment because of RNA's previously proven ability to bind strongly to it.
- Transcription- the process of producing an RNA copy of one strand of a DNA template.